

*До 100-річчя  
Національної академії наук України*

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ  
УКРАЇНСЬКИЙ МОВНО-ІНФОРМАЦІЙНИЙ ФОНД

# ЛІНГВІСТИЧНО-ІНФОРМАЦІЙНІ СТУДІЇ

ПРАЦІ  
УКРАЇНСЬКОГО МОВНО-ІНФОРМАЦІЙНОГО ФОНДУ  
НАН УКРАЇНИ

У П'ЯТИ ТОМАХ

Дослідження 1991–2018 років

За загальною редакцією *В. А. Широкова*

КИЇВ 2018

НАЦІОНАЛЬНА АКАДЕМІЯ НАУК УКРАЇНИ  
УКРАЇНСЬКИЙ МОВНО-ІНФОРМАЦІЙНИЙ ФОНД

ТОМ III

**ТЛУМАЧНА ЛЕКСИКОГРАФІЯ**

КНИГА 1

**СЛОВНИК УКРАЇНСЬКОЇ МОВИ  
У ДВАДЦЯТИ ТОМАХ**

УКРАЇНСЬКИЙ МОВНО-ІНФОРМАЦІЙНИЙ ФОНД

*Рекомендовано до друку вченою радою  
Українського мовно-інформаційного фонду НАН України  
(протокол № 9 від 23 жовтня 2018 р.)*

Рецензенти:

академік НАН України *О. В. Палагін*,  
доктор філологічних наук, професор *А. А. Лучик*

Автори першої книги третього тому:

*В. А. Широков, Н. М. Заїка, А. П. Загнітко, Г. В. Карнаух, І. М. Кондрачук, М. Ю. Кригін,  
Л. М. Кудрявкіна, Ю. В. Леонтєв, В. В. Манак, М. В. Надутенко, Л. В. Оліфіренко,  
І. В. Остапова, О. Г. Рабулець, Л. М. Томіленко, В. В. Сивокозова, Л. О. Симоненко, В. В. Чумак,  
І. В. Шевченко, Л. Л. Шевченко, Н. В. Щербакова, К. М. Якименко, Г. М. Ярун*

**Лінгвістично-інформаційні студії**: Праці Українського мовно-інформаційного фонду НАН України : у 5 т. Т. 3 : Тлумачна лексикографія. Кн. 1 : Словник української мови у двадцяти томах / В. А. Широков та ін. Київ : Український мовно-інформаційний фонд НАН України, 2018. 276 с.

Першу книгу «Словник української мови у 20 томах» третього тому «Тлумачна лексикографія» п'ятитомної монографії «Лінгвістично-інформаційні студії» присвячено викладу лінгвістичних і технологічних засад сучасної тлумачної лексикографії. Його виконано на прикладі проекту «Словник української мови у 20 томах», який здійснюється Українським мовно-інформаційним фондом.

Концептуальну основу роботи становлять теорія лексикографічних систем і теорія семантичних станів, викладені у першому томі монографічного видання. Зокрема, тут сформульовано загальний вигляд семантичного стану для одиниць лексичного рівня мови, здійснено декомпозицію цього стану на лексико-граматичну та лексико-семантичну компоненти й ураховано взаємодію граматичної та лексичної семантики, а також параметри синтактики і прагматики. Розроблено структуру лексикографічної системи «Словника української мови», експліковано елементи метамови цієї структури та їх маркери в тексті Словника.

Сформульовано основні принципи лексикографування «Словника української мови у 20 томах» (СУМ-20) як сучасного фундаментального академічного твору, який має загальнонаціональне значення. Розроблено критерії добору лексики до Словника. Докладно відпрацьовано й викладено правила подання граматичної інформації до одиниць реєстру в розрізі окремих лексико-граматичних класів. Особливу увагу приділено поданню семантичної інформації і насамперед – розробленню словникових дефініцій. Останні деталізуються як за частиномовною належністю відповідних реєстрових одиниць, так і за їх належністю до певних лексико-семантичних груп. Значну увагу приділено принципам і правилам добору текстово-ілюстративного матеріалу до семантичних станів лексичних одиниць.

Оскільки принциповою відмінністю від інших словникових проектів стала орієнтація проекту СУМ-20 на комп'ютерно-комунікаційні технології, то велику увагу в монографії приділено викладу інноваційної інформаційно-комунікаційної технології створення і розвитку тлумачних словників у формі віртуальної лексикографічної лабораторії (ВЛЛ СУМ). З цією метою описано базисні поняття, стандарти та специфікації практичної реалізації ВЛЛ СУМ, пов'язані з застосуванням системотехніки сервіс-орієнтованої архітектури. Формально визначено поняття лексикографічного процесу та подано методичку його реалізації у ВЛЛ СУМ. Здійснено виклад процесу парсингу СУМ-11 та автоматичної побудови на цій основі відповідної лексикографічної бази даних, яка стала основою для формування ВЛЛ СУМ. Розроблено користувацький інтерфейс ВЛЛ СУМ (зовнішню модель відповідної Л-системи), описано функції його елементів і технологію роботи з ними; сформульовано принципи адміністрування лексикографічним процесом. Викладено принципи лексикографічної статистики та результати їх застосування для порівняння динаміки лексикографічних параметрів від СУМ-11 до СУМ-20

# ЗМІСТ

ВСТУП. ТЛУМАЧНА ЛЕКСИКОГРАФІЯ НА ЗАГАЛЬНОЛІНГВІСТИЧНОМУ ТЛІ .....	7
Загальні положення .....	7
Означення, функції та загальна структура словника.....	11
<b>Розділ 12. ЛІНГВІСТИЧНІ ЗАСАДИ СУЧАСНОЇ УКРАЇНСЬКОЇ ТЛУМАЧНОЇ ЛЕКСИКОГРАФІЇ .....</b>	<b>19</b>
12.1. Тлумачний Словник української мови у 20 томах .....	19
12.2. Граматична та лексична семантика в тлумачному словнику: традиційний та формалізований підходи .....	24
12.2.1. Лексичне значення слова та його лексикографічне представлення .....	24
12.2.2. Формалізація значень у тлумачному словнику .....	35
12.3. Головні принципи будови та укладання СУМ-20 .....	41
12.3.1. Будова лексикографічної системи Словника української мови .....	41
12.3.2. Структура тексту СУМ-20, його метамова та її маркери в тексті Словника .....	49
12.3.3. Основні принципи лексикографування .....	53
12.3.4. Лексичний склад Словника. Критерії добору лексики та формування реєстру .....	58
12.3.5. Орфографія СУМ-20.....	61
12.3.6. Граматичні форми і граматична характеристика реєстрових слів.....	67
Уваги до подавання окремих частин мови .....	70
А. Іменник .....	70
Б. Прикметник.....	75
В. Числівник.....	76
Г. Займенник .....	76
Ґ. Дієслово .....	77
Д. Дієприкметник .....	85
Е. Прислівник.....	86
Є. Прийменник.....	87
Ж. Сполучник .....	87
З. Частка.....	87
И. Вигук.....	87
12.3.7. Ремаркування реєстрових слів і фразеологізмів .....	88
12.3.8. Подання тлумачень. Загальні положення .....	97
12.3.9. Тлумачення реєстрових слів за їх належністю до частини мови .....	100
А. Іменник .....	100
Б. Прикметник.....	101
В. Числівник.....	102
Г. Займенник .....	103
Ґ. Дієслово .....	103
Д. Дієприкметник .....	104
Е. Прислівник.....	105
Є. Прийменник.....	106
Ж. Сполучник .....	106
З. Частка .....	107
И. Вигук.....	108

12.3.10. Тлумачення реєстрових слів за їх належністю до певних лексико-семантичних груп. Енциклопедичний елемент у тлумаченні .....	108
12.3.10.1. Тлумачення реєстрових одиниць – назв рослин .....	109
12.3.10.2. Тлумачення реєстрових одиниць – назв тварин.....	110
12.3.10.3. Іменники – назви викопних тварин і рослин.....	111
12.3.10.4. Іменники – назви наукових дисциплін.....	111
12.3.10.5. Тлумачення іменників – назв фізичних одиниць виміру .....	112
12.3.10.6. Тлумачення іменників – назв хімічних елементів .....	112
12.3.10.7. Тлумачення іменників – назв мінералів.....	113
12.3.11. Подання словосполучень .....	113
12.3.12. Текстово-ілюстративний матеріал .....	122
12.3.13. Індокси СУМ-20 .....	128
12.3.14. Поліграфічне оформлення СУМ-20 .....	131
12.3.15. Особливості редакційної підготовки Словника української мови у 20 томах.....	147

### **Розділ 13. ТЕХНОЛОГІЧНІ ІНСТРУМЕНТИ ПРОЕКТУ «СЛОВНИК УКРАЇНСЬКОЇ МОВИ» .....**

<b>13.1. Віртуальна лексикографічна лабораторія «СУМ» .....</b>	<b>158</b>
13.2. Лінгвістичний корпус.....	179
13.3. Функціонал і користувацький інтерфейс віртуальної лексикографічної лабораторії «Словник української мови» .....	179
13.3.1. Лексикографічний процес .....	179
13.3.2. Парсинг .....	185
13.3.3. Підхід до проектування користувацького інтерфейсу .....	190
13.3.4. Адміністрування доступу.....	192
13.3.5. Редагування лексикографічних об'єктів.....	195
13.3.6. Пошук лексикографічних об'єктів .....	214
13.3.7. Комунікації з лексикографічними об'єктами.....	219
13.3.8. Моніторинг лексикографічних об'єктів .....	223
13.3.9. Синхронізація текстів технологічних томів .....	226
13.3.10. Успадкування текстів технологічних томів ЛС СУМ.....	232
13.3.11. Маркування та індексування лексикографічних об'єктів.....	233
13.3.12. Тестування лексикографічних об'єктів .....	237
13.3.13. Лексикографічна статистика.....	248
13.3.14. «Словник української мови» on-line .....	253
<b>ВИСНОВКИ.....</b>	<b>256</b>
<b>ДОДАТОК. Таблиця ранжування джерел СУМ-20 .....</b>	<b>257</b>
<b>ЛІТЕРАТУРА.....</b>	<b>273</b>

# ВСТУП

## ТЛУМАЧНА ЛЕКСИКОГРАФІЯ НА ЗАГАЛЬНОЛІНГВІСТИЧНОМУ ТЛІ

### Загальні положення

На початку ХХІ століття лексикографія постала як напрям людської діяльності, що, з одного боку, осмислив і втілив у творах словникарства багатовіковий досвід теоретичних надбань лінгвістики й практичного лексикографування, а з іншого – отримав досі небачені перспективи розвитку, пов'язані із застосуванням комп'ютерних технологій. Відбувається неухильне зростання соціальної значущості лексикографічної продукції. Сучасна наука прагне втілити у словниковій формі всі аспекти отриманих знань; словники не лише фіксують сукупність певних знань, але й є надійним інструментом наукового освоєння дійсності. Словникова форма як тип соціальної комунікації все більше приваблює спеціалістів найрізноманітніших предметних галузей, оскільки «за повнотою, змістовністю, точністю й достовірністю інформації словники, безперечно, займають перше місце серед інших форм, які представляють дані у стислому вигляді»<sup>1</sup>. Не випадково підсумки лінгвістичних робіт нерідко втілюються у словниках. Ю. М. Караулов пише про це так: «пафос опису будь-якого рівня або фрагмента мови вбачається у створенні словника, його елементів, процедур, відношень»<sup>2</sup>.

Відповідно до потреб сьогодення з'являються нові розділи лексикографії, які подають приклади далекосяжних узагальнень поняття словника, а інтенсифікація інформаційних процесів спонукає до прискорення розроблення різноманітних методик, пов'язаних з формалізацією мови, і до створення все потужніших методів лексикографування явищ предметного світу. Шляхи, якими прямує новітня лексикографія, багато в чому визначаються зовнішніми щодо власних завдань лінгвістики чинниками, серед яких найголовнішим є розширення масштабів науково-технічної революції у галузі інформаційних систем та комп'ютерних комунікацій, а також глобалізація, становлення індустрії знань і викликана цим потреба в інтелектуальних засобах екстракції знань (knowledge mining), здатних у реальному часі опрацьовувати надвеликі масиви природномовної інформації. Тому закономірно, що словникова справа сьогодні переживає особливий етап свого розвитку, перебуваючи під великим впливом

---

<sup>1</sup> Герд А. С. К определению понятия «словарь». Проблемы лексикографии: сб. статей. Санкт-Петербург : Изд-во СПбГУ, 1977. С. 198.

<sup>2</sup> Караулов Ю. Н. Лингвистическое конструирование и тезаурус литературного языка. Москва : Наука, 1981. С. 42.

нових суспільних потреб і нових методів оброблення інформації, використовуючи можливості в застосуванні комп'ютерних технологій при описі та представленні як власне лінгвістичної, так і екстралінгвістичної інформації. Отже, можна констатувати, що новітня комп'ютерна лексикографія являє собою: а) прикладну наукову дисципліну на межі мовознавства та інформатики, яка вивчає способи і прийоми застосування методів інформаційної науки й технології в теорії та практиці створення широкого спектра лексикографічних систем; б) галузь комп'ютерної індустрії, яка швидко розвивається головним чином через те, що лексикографування знань є одним з ефективних способів їх виявлення і поширення. Автори, однак, свідомі того, що навіть сам термін «комп'ютерна лексикографія» несе в собі рефлексію перехідного періоду розвитку мовознавства, коли нові інформаційні технології ще не стали універсальним інструментом мовознавства, а лінгвістичні дані (тексти, лексикографічні структури та ін.) побутують паралельно в обох формах – як традиційній, так і цифровій. Цілком можливо, що з часом уся лексикографія стане комп'ютерною і потреба в окремому терміні, який вирізняє її від традиційної, відпаде.

Оскільки комп'ютерна лексикографія (як і будь-яка комп'ютерно-орієнтована галузь знання) вимагає максимальної формалізації свого об'єкта, виникла потреба в поглибленні змісту і формалізації самого поняття словника як специфічного мовно-інформаційного об'єкта, в який «прагне» втілитися мовна субстанція у процесі розгортання фундаментальних для мови відношень «суб'єкт – об'єкт» і «форма – зміст». Відповіддю на цю потребу стало розроблення одним з авторів цієї книги теорії лексикографічних систем<sup>3</sup>, яка на основі досить загальних припущень і положень теорії інформації та ґрунтуючись на знов-таки найзагальніших властивостях вищезгаданих системотвірних для мови відношень, дала змогу пояснити внутрішні механізми, що спонукають до набуття природномовною (і не тільки) інформацією словникової форми.

Обговорюючи лексикографічну проблематику, не можна не згадати принаймні чотири важливі проблеми, які особливо актуалізувалися в наш час.

*Першою з них є проблема оновлення фундаментальних лексиконів.* Адже великі словники, звичайно, є взірцевими лексикографічними об'єктами, створеними значними колективами висококваліфікованих фахівців. Ці об'єкти є яскравими зразками лінгвістичної думки певного історичного періоду і несуть велику кількість явно і неявно виражених рефлексій з тією чи іншою епохою, у них «вкарбована» історична пам'ять народу та його мовна свідомість у певний період. Поряд із цим такі лексикографічні твори мають інваріантну цінність, виражену в їхніх лексикографічних структурах, і подають новим генераціям лексикографів нетлінні зразки способів лексикографування семантичних явищ, приклади розв'язання лексикографічними засобами доволі складних лінгвістично-когнітивних завдань. З огляду на сказане такі лексикографічні твори при

---

<sup>3</sup> Широков В. А. Інформаційна теорія лексикографічних систем. Київ : Довіра, 1998. 331 с.



завершенні певної епохи не варто здавати в архів, але доцільно повною мірою використати найцінніше, що в них є. Такий погляд веде до розв'язання непростих завдань щодо оновлення фундаментальних лексикографічних праць, модифікування їх змісту і структури та приведення їх у відповідність з новими реаліями. І тут виникає ціла низка конкретних завдань – починаючи від визначення лінгвістичних і когнітивних принципів укладання нового лексикону «поверх» старого й закінчуючи технологією та організацією виконання такого лексикографічного проекту. При цьому технологічні проблеми дедалі виразніше пересуваються на передній план актуальності, що пов'язано насамперед зі значним стисненням термінів, які відводяться на модернізацію лексикона й необхідністю подання кінцевого результату виконаної роботи в різних, «неізоморфних» формах (традиційній книжковій, у вигляді лазерного диска, сайту в інтернеті, резидентного словника в комп'ютерних системах опрацювання текстів, електронного семантичного аналізатора та ін.).

*Друга проблема* пов'язана з необхідністю переведення лексикографічної спадщини людства у цифрову форму для нагромадження потужного загальнокультурного потенціалу, а також для використання в системах автоматичного опрацювання інформації. Для цього абсолютно недостатньо просто відсканувати тексти відповідних словників. Хоча й це не таке просте завдання, як може показатися на перший погляд. Відомо, що метаслова багатьох словників містить низку нестандартних символів, які важко піддаються ідентифікації. Крім того, існуючі програми оптичного розпізнавання текстів далеко небездоганні й припускаються багатьох помилок у процесі їхнього застосування. Нарешті, робота зі словниками, переведеними у комп'ютерні текстові формати, є дуже неефективною і потребує конвертування їх у формати лексикографічних баз даних, що є спеціальним завданням, не відомим класичній лексикографії.

*Третя проблема* полягає в застосуванні словників для формування лінгвістичних компонентів концептографічних систем представлення знань (типу онтології) та використання їх у засобах knowledge mining. Це потребує не лише великих словникових масивів, але й спонукає до віднаходження у традиційних словникових текстах прихованих семантичних структур. З іншого боку, онто- і концептографічна проблематика спонукає до більш активної і змістовної інтеграції лексикографії до сучасної індустрії знань.

*Четверта проблема* стосується подолання «вавилонського синдрому», тобто вирішення значного кола питань щодо багатомовності, які особливо загострилися в епоху глобалізації і дедалі наполегливіше вимагають потужних і гнучких лексикографічних засобів для свого розв'язання. Серед останніх особливо відзначимо інструментальні засоби лексикографії, налаштовані на комп'ютерно-технологічну підтримку процесів багатомовного лексикографування

в різних режимах, зокрема й віддаленому, який забезпечує творчу взаємодію територіально розподілених інтернаціональних колективів лексикографів.

Зрозуміло, що вказана проблематика вимагає від представників комп'ютерної лексикографії певної суми знань та умінь, нетипових і незвичних для попереднього періоду розвитку лексикографічної науки й практики. Серед таких відзначимо здатність до формалізації мовного матеріалу, доведення його до формалізованих моделей даних, схильність до системно-структурного мислення, хоча б елементарні знання про будову комп'ютера, уміння працювати з базами даних, навички програмування (принаймні на рівні формування SQL-запитів) і т. ін.

Ця книга покликана передусім продемонструвати на прикладі фундаментальних тлумачних лексиконів, як вони можуть бути використані для розв'язання вищезазначених проблем. Адже у монографічній літературі майже немає зразків комп'ютерного укладання словника від початку і до кінця – починаючи від задуму, побудови теоретичної моделі лексикографічної системи словника, і закінчуючи описом процесу лексикографування й аналізом одержаних результатів.

Робота зі Словником української мови дала змогу не тільки глибше усвідомити лінгвістичні ефекти, «зашифровані» в структурі Словника, але й створити підґрунтя для побудови в автоматичному режимі вельми розвиненої і складноструктурованої комп'ютерної бази даних Словника української мови. Наскільки нам відомо, досі у світовій комп'ютерній лексикографії не вдавалося здійснити аналогічну роботу для лексиконів такого масштабу і складності.

Особливо зазначимо, що для реалізації проекту «Словник української мови» була розроблена спеціальна комп'ютерна технологія, яка ґрунтується на методі віртуальних лексикографічних лабораторій (ВЛЛ). Ця технологія надає засоби для виконання спільних лексикографічних проектів фахівцями різних територіально розподілених інституцій і навіть різних країн. ВЛЛ «СУМ» розробляється на базі Web-сервісів (WS) – однієї з реалізацій SOA-дodatка, що позиціонується як універсальна технологія зв'язування суттєво різнорідних систем. У цій технології використовуються кілька базисних стандартів: XML для опису даних, SOAP для передання інформації між системами, WSDL для опису сервісів та UDDI для збереження й надання WSDL-описів за запитом. Цих стандартів досить для створення відносно нескладних систем. Однак мінімально нетривіальні рішення, що зазвичай і потрібні в корпоративному середовищі, вимагають забезпечення таких аспектів, як гарантоване асинхронне доставлення повідомлень, керування транзакціями, шифрування даних, які циркулюють між системами, і забезпечення їх недоторканності. Ці аспекти так чи інакше пов'язані з WS і тепер активно розробляються інструменти надбудови різних специфікацій.

Стосовно призначення цієї книги, зауважимо, що вона адресована не лише практичним словникарям, яким корисно ознайомитися із сучасними лінгвістичними концепціями та технологічними підходами до реалізації

фундаментальних лексикографічних проєктів. Важливим адресатом монографії, на нашу думку, є й широке коло мовознавців та інформатиків – як теоретиків, так і представників прикладної лінгвістики, лінгвістичної технології, когнітивістики та інших галузей науки, хто зацікавлений у фундаментальних інформаційно-логічних підвалинах лексикографії та усвідомлює її зв'язки із різними галузями лінгвістики, логіки та інформатики. Авторам було важливо показати інтелектуальний і поняттєвий діапазон сучасної лексикографічної науки, продемонструвавши її зв'язок з математичною логікою, теорією моделей даних, інтегрованих інформаційних систем і логіко-лінгвістичних числень. Цій меті сприяє й виклад у даному томі неklasичних підходів до лексикографічної проблематики, серед яких особливо відзначимо концепцію семантичних станів і лексикографічних числень. Звертаємо також увагу на вперше висвітлені в монографічній літературі зв'язки лексикографічних процесів і структур з такими фундаментальними концептуальними результатами математичної логіки, як теорема Левенгейма-Сколема та теорема Геделя. Вважаємо, що монографія буде корисною також студентам і аспірантам спеціальності «Структурна, прикладна та математична лінгвістика», оскільки у вітчизняній науковій літературі дуже відчувається брак підручників та посібників, присвячених проблемам комп'ютерної лексикографії, яка є одним з основних розділів згаданої спеціальності.

## **Означення, функції та загальна структура словника**

Для ефективного розв'язання поставлених проблем необхідно мати певний узагальнений погляд на словник як на об'єкт, що є втіленням певних аспектів лексикографічного опису мовної системи. Тому необхідна загальна формальна теорія словників – як варіант такої теорії нами було побудовано теорію лексикографічних систем. Крім цієї загальної формалізованої схеми необхідною вважаємо побудову певних «містків» – змістових інтерпретацій цієї схеми на класичних лексикографічних об'єктах, якими насамперед є великі тлумачні лексикони національних мов.

Узагалі філологічна наука за багатовікову історію свого існування й розвитку виробила системні підходи до означення поняття словника. У вузькому розумінні словник (лексикографічний твір) – це сукупність мовних одиниць, розташованих у певному порядку, де розкрито їхнє значення, подано про них різні відомості чи переклад іншою мовою або вміщено інформацію про предмети, явища та факти, які вони позначають.